METHOD FOR CONVERTING HIGH LEVEL MOTION SCRIPTS TO COMPUTER ANIMATIONS

BACKGROUND OF THE INVENTION

Field of Invention

The invention is related to a 3D (standing for three-dimensional) animation generation method used in digital multimedia, especially related to a 3D animation generation method using high-level motion scripts.

Related Art

In recent years, the application areas of computers have been broadened by their increasing computation power. With the advance of digital multimedia techniques, mass media also use computers to produce and deliver contents. In addition, recreation companies have already employed computer-based techniques to create animations and synthesize virtual characters in computer games. How to generate vivid and controllable character animations becomes an important issue in the areas of computer animation and video games.

In the traditional animation production, the motions of each character are drawn frame by frame by animators. Even for keyframes, describing a pose requires setting the angles of all joints, and hence requires setting about 20 to 60 parameters for each frame. As a result, it is difficult to animate and control virtual characters on the fly. Besides, the keyframe method heavily relies on animators' skills and experiences to produce vivid human animations. Another approach is known as the kinematics-based animation production method. When creating human animations, the method calculates the translation and rotation parameters of

the end-effectors, the angles of joints, centers of gravity and roots by using techniques of biomechanics to generate vivid animations. Due to the high complexity of human motions, it is difficult to find good approximate motion equations. Hence, the application of this method is restricted, and is usually used in the syntheses of locomotion animations.

Dynamics is another method for simulating and generating motions by formulating the mass, inertia and angular moment of objects. However, simulating complicated joint systems such as human beings consumes a lot of computation power. Hence, it is difficult to generate animations by real-time dynamic simulation. The latest method employs 3D motion sensors to capture human motions. Since the captured motion data are guaranteed to fulfill the constraints in dynamics, the captured motion data are more vivid than those obtained by the prior methods. However, motion capture equipments are expensive and both capture and data editing processes are time-consuming. To reduce these costs, the reuse of the captured motion data becomes an important research issue. Recently, motion graphs and motion texture proposed novel control mechanisms to synthesize a new motion based on the existing motion data. However, these approaches still remain some difficulties such as long preprocessing time, and unexpected transitions. Moreover, the connection between high-level motion control and low-level mathematical models developed by these systems is unclear.

SUMMARY OF THE INVENTION

To solve the mentioned problems, the invention proposes a 3D animation generation method, which enables users to synthesize 3D animations by inputting natural language scripts.

The invention is related to a 3D animation generation method using scripts to automatically synthesize 3D animations by natural language analysis techniques and the motion index tables. In essence, the proposed method is able to generate various 3D animations by using an annotated human motion database and the natural language analysis techniques. The proposed method first analyzes the motion- related terms and knowledge in natural language processing, and builds their ontology. Then, the ontology is transformed into semantic metadata to enable computers to understand the semantics of natural language. Finally, the required motion clips are retrieved from the motion database, and are synthesized into a 3D animation.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the control flow of the proposed method.

Figure 2 shows the control flow of the proposed natural language formalization method.

Figure 3 shows the control flow of the establishment of a motion database.

Figure 4 shows the control flow of the proposed motion clip search method.

DETAILED DESCRIPTION OF THE INVENTION

The invention proposes a 3D animation conversion method using scripts. Figure 1 shows the control flow of the proposed method. After receiving a user-inputted high-level motion script (Step 101), the method first formalizes the script into a computer-recognizable formation (Step 102), then compares the formalized script with the annotation in the motion database (Step 103), retrieves the corresponding motion clips (Step 104), and finally,

synthesizes these motion clips into a 3D animation (Step 105).

Formalizing natural language into a computer-recognizable formation is the foundation of the proposed method. Hence, we take thesauruses and metadata to perform formalization. Figure 2 shows the control flow of the formalization of natural language. First, we apply part of speech tagging to the natural language script (Step 201). Then, the part of speech (Step 202) and the corresponding formal representative of each word are recognized (Step 203). Accordingly, we form the formal script according to formal constructs (Step 204). Since the script is composed of natural language terms, transforming the script from natural language into the formal language relies on the thesauruses, which are used to keep the consistency of metadata and to store the mapping of the terms with similar meanings in the specific domain. Since natural language is not annotated by any semantic metadata, computers cannot understand the high level semantics of the natural language in the digital content. Hence, metadata annotation is used to enable computers to understand the implicit semantics of the digital content. However, metadata must be well formed. This criterion enables users to annotate the semantics of digital content under some guidelines, and enables computers with limited ontology and inference rules to understand human's thoughts and creativity.

Take a human body animation as an example. Since human motions can be expressed by specific terms, the thesauruses are established to generate the mapping of metadata. First, human motions related documents are collected and analyzed by natural language processing tools (also known as natural language parsers) to tag the part of speech of each word in the documents (e.g., noun, verb, preposition...). According to the statistics of these tags, keywords are extracted and thesauruses are built. Then, we use thesauruses to map the synonyms of these keywords into formal representatives. For example, "move downward" is used as the formal representative of "downward", "move down" and "go down".

Accordingly, the motion data can be annotated by metadata. Metadata can be expressed in XML (standing for Extensible Markup Language) format to obtain portability and generality.

After formalization, a formalized script is formed and used to compare with the annotations, which are also formalized scripts, in the motion database to retrieve the corresponding motion clips to synthesize a 3D animation. The motion database comprises several motion clips and motion index tables. The corresponding motion clips can be retrieved by using the motion index table and comparing the metadata of corresponding motion clips. Figure 3 shows the control flow of the establishment of the motion database. First, the motion data are read (Step 301). Then, the coordinates of each frame in a motion data are extracted (Step 302) and the coordinate features are calculated (Step 303). The motion clips and the corresponding index table are established according to the coordinate features (Step 304). In the motion capture data, each frame records the 3D Cartesian coordinate of each joint and the root orientation. Take the human motion as an example. In each frame, we first extract the poses of the limbs (i.e., left arm, right arm, left foot and right foot). An arm comprises an upper arm and a forearm, and a foot comprises a thigh and a calf. In order to reduce the number of dimensionality and to be affine invariant during body movement, the representation of the limb is transformed from their 3D Cartesian coordinates to 2D spherical coordinates. Let v be a limb vector and r be equivalent to the root orientation vector. Suppose π is the plane passing through the joint o and parallel to the floor. Let the projection of v and r on π be the v_{xz} and r_{xz} respectively. Then θ and φ, the spherical coordinates of v on π, are measured in angular radians from v_{xz} to r_{xz} and from v to Y axis respectively. In this case, Y axis is the normal vector of π.

An arm posture is represented as 4D tuples $(\theta, \varphi, \hat{\theta}, \hat{\varphi})$, where (θ, φ) and $(\hat{\theta}, \hat{\varphi})$ are

extracted from the upper arm and the forearm, respectively. We also use the same steps to extract the features of a foot.

As shown in Figure 4, a motion index table is a direct sum of two four-dimensional index tables (the poses of the left and right arms) and an eight-dimensional index table (the poses of both feet). For each motion frame, we quantize its posture features to form its index. For example, $(\theta, \varphi, \hat{\theta}, \hat{\varphi})$ are the posture features of the left arm in the i-th frame, then its index can be computed using the following truncation function H,

$$H(\theta_i, \varphi_i, \hat{\theta}_i, \hat{\varphi}_i) = \left(\left\lfloor \frac{\theta_i}{a} \right\rfloor, \left\lfloor \frac{\varphi_i}{b} \right\rfloor, \left\lfloor \frac{\hat{\theta}_i}{c} \right\rfloor, \left\lfloor \frac{\hat{\varphi}_i}{d} \right\rfloor \right),$$

where a, b, c, d are the step sizes of angle radians and the operator $\lfloor \ \rfloor$ denotes the floor function. A set of successive frames will be indexed into the same cell by the above equation as long as they are with the same truncated posture features. Hence, the successive motion captured data will be partitioned into several consecutive cells, and each cell may contain several motion clips. The numbers of the starting and ending frames in each motion clip are also stored in the corresponding cell.

The motion index table can be established when all motion data have been partitioned well. As shown in Figure 4, the motion index table 40 contains multiple cells, which comprise several cells with data 401 and several cells without data 402. The more motion data there are, the less the number of cells without data is, and the less the restriction of generating animation is. In addition, the metadata of motion annotations have to be consistent with the indexed motion data. The hierarchical MPEG-7 DDL (standing for description definition language) format is then used to annotate the motion features and the semantics of

the static and dynamic motions. A normal form similarity matching mechanism is applied to approximate the best matching between the formalized input script and the annotation of the motion database. The continuous DTW (standing for dynamic time warping) algorithm is employed. The corresponding cells of the pose and the corresponding cell connection path of the motion can be obtained according to the similarity of the metadata. Finally, the information of the starting and ending frames can be obtained from the metadata of the pose and the motion.

Figure 4 shows the steps from pose indexing to motion synthesis. Pose indexing is to find the cells of the starting frame A and ending frame B. Suppose that the starting and ending frames are f_{start} and f_{end}, respectively, and the corresponding cells are C_{start} and C_{end}, respectively. Path searching is to find the possible paths from C_{start} to C_{end}. For example, there are three possible paths in Figure 4. The path is determined by an algorithm which uses a threshold α to restrict the search space, and adjusts the weights according to the numbers of motion clips in a cell. This algorithm repeats until a path from C_{start} and C_{end} is discovered. After the assignment of all key poses, the system retrieves the corresponding motion capture data according to the motion index table, and obtains the connection paths of key poses by visiting neighbor cells with a greedy algorithm. The motion transitions among neighbors should not only consider the root orientation and the alignments of the motion directions, but also solve the feet sliding, penetrating, suspending on the floor, and other phenomena violating environmental constraints.

While the preferred embodiment of the invention has been set forth for the purpose of disclosure, modifications of the disclosed embodiment of the invention as well as other embodiments thereof may occur to those skilled in the art. Accordingly, the appended claims are intended to cover all embodiments, which do not depart from the spirit and scope of the

invention.